Ein einfaches Programm zur Berechnung von gewogenen gleitenden Mittelwerten in dBASE

Rahlf, Thomas

Historical Social Research, Vol. 18 — 1993 — No. 1, 122-144

METHODS: REVIEWS AND NOTICES

Ein einfaches Programm zur Berechnung von gewogenen gleitenden Mittelwerten in dBASE

*Thomas Rahlf**

Abstract: The application of moving averages in the course of investigation on historical time series has certain disadvantages. Different weighing the single elements can provide a considerable improvement. The following article tackles the matter of weighing and its effects, explaining three different methods: weighing by local trendpolynoms of the 3^{rd} and 5^{th} degree, weighing on the basis of the binomial- and normaldistribution. Finally the calculation of these weights and weighted moving averages on the basis of a problem-oriented programming language is shown.

Die Anwendung von gleitenden Mittelwerten bei Zeitreihen ist ein auch von Historikern oft benutztes Verfahren. (1) Damit werden vor allem zwei Ziele verfolgt: (2)

I. Zufällige Schwankungen sollen eliminiert werden. Betrachtet man eine Zeitreihe als Zusammensetzung verschiedener -zyklischerSchwankungen, benutzt man einen gleitenden Mittelwert, um Schwankungen einer bestimmten Länge hervortreten zu lassen und andere zu eliminieren. Die Zeitreihe wird also geglättet.
II. Auf der anderen Seite werden gleitende Mittelwerte verwendet, um periodische Schwankungen einer Zeitreihe zu eliminieren: bekanntestes Beispiel sind etwa Saisonschwankungen.

Die Vorgehensweise ist für beide Ziele gleich; lediglich die Wahl des Stützbereiches (3) entscheidet darüber, ob bestimmte Zyklen einer Reihe eliminiert oder hervorgehoben werden. Enthält eine monatliche Zeitreihe etwa saisonale

---

* Address all communications to Thomas Rahlf, Stiftsgasse 2, 5300 Bonn 1. Eine ohne dBASE ablauffähige Version, die vollständig menügeführt ist, fehlende Werte berücksichtigt und sowohl eine ASCII-, als auch eine DBF-Datei erzeugt, kann zu public-domain Bedingungen bezogen werden.

Schwankungen, so werden diese bei der Wahl eines Stützbereiches von M = 12 eliminiert, bei M = 6 dagegen betont. Leider ist die Anwendung von gleitenden Mittelwerten mit Problemen verbunden:

- In der Realität ist die Länge der Zyklen in der Regel nicht von gleicher Länge. Durch die Anwendung des Mittelwertes können dann Verschiebungen der Spitzen und Täler entstehen.
- Im Extremfall können durch einen gleitenden Mittelwert sogar künstliche Zyklen erzeugt werden: der sogenannte »Slutzky-Effekt«. (4)

Die Abbildungen 1 bis 4 zeigen eine hypothetische Reihe mit zwei siebenjährigen und knapp zwei dreizehnjährigen Zyklen. Für diese Reihe wurden gleitende Mittelwerte der Länge M = 7, M = 11 und M = 13 berechnet. Wählt man für diese Reihe einen Stützbereich von M = 7, so werden die ersten beiden Zyklen exakt eliminiert (Ziel II), die zweiten dagegen nahezu vollständig wiedergegeben (Ziel I). Wählt man einen Stützbereich von M = 11, werden durch den Mittelwert die ersten beiden Zyklen umgekehrt, die letzten beiden dagegen nicht. Bei einem Stützbereich von M = 13 werden die letzten beiden Zyklen vollständig eliminiert, dagegen das erste und dritte Tal sowie die zweite Spitze umgedreht. Da gerade in historischen Zeitreihen mit einer Vielzahl von unterschiedlich langen Zyklen zu rechnen ist, leuchtet ein, daß die Anwendung herkömmlicher gleitender Mittelwerte für ihre Untersuchung ungeeignet ist.

Ein sehr anschauliches Mittel, die Eigenschaften von gleitenden Mittelwerten zu untersuchen, ist das sog. Amplitudendiagramm. Man betrachtet hierzu den gleitenden Mittelwert als »Filter«, der je nach Stützbereich bestimmte Übertragungseigenschaften aufweist. Auf den mathematischen Hintergrund wird hier nicht näher eingegangen, dafür sei auf weiterführende Literatur verwiesen. (5) Idealerweise sollte ein solcher Filter genau die Schwingungen einer Zeitreihe entweder eliminieren oder herausfiltern, an denen man interessiert ist. Ein Filter, der etwa alle Schwankungen bis zu einer Länge von sieben Jahren in die geglättete Reihe passieren lassen soll, müßte demnach folgende Gestalt haben: Alle Schwankungen der Zeitreihe die länger oder gleich sieben Zeiteinheiten sind, werden vollständig in die gefilterte Reihe übertragen (G(co) = 1), alle Schwankungen, die kürzer sind, vollständig eliminiert (G(co) = 0, vgl. Abb. 5). Betrachtet man das Amplitudendiagramm eines siebengliedrigen gleitenden Mittelwertes, so zeigt sich der in Abb. 6 dargestellte Verlauf. Die Funktion hat zwar bei der Frequenz, die sieben Zeiteinheiten entspricht, ihre erste Nullstelle, bis dahin jedoch keinen rechteckigen, sondern einen kontinuierlich sinkenden Verlauf. Anschließend erscheinen sog. »side-lobes« mit Spitzen bei ca. 4.87, 2.83 und 2 Zeiteinheiten. Betrachtet man die Fläche unterhalb des Graphen, so beträgt die Fläche unter diesen »side-lobes« ca. 1/3 der Gesamtfläche. Solch ein Amplitudendiagramm ist typisch für gleitende Mittelwerte dieser Art: mit zunehmender Größe des Stützbereiches erhöht sich auch die Anzahl der Fläche rechts von der ersten Nullstelle. (6)

Welche weiteren Möglichkeiten ergeben sich nun? Eine mögliche Antwort bietet die Konstruktion von digitalen Filtern, indem man von idealen Transferfunktionen ausgeht. Solch ein Vorgehen erfordert jedoch ein hohes mathematisches Abstraktionsniveau und dürfte den meisten Historikern, deren Ausbildung nicht unbedingt mathematisch orientiert ist, wie eine »black box« erscheinen (7).

Demgegenüber haben gleitende Mittelwerte den Vorteil, daß ihre Berechnung intuitiv verständlich ist. Im folgenden wird deshalb gezeigt, wie sich die Eigenschaften von gleitenden Mittelwerten durch Gewichtungen der einzelnen Elemente erheblich verbessern lassen.

Solche Berechnungen erweisen sich von Hand als sehr aufwendig. Leider ist auch ihre Umsetzung in Standard-Anwendungsprogrammen nicht vorgesehen oder nur sehr umständlich zu realisieren.

Im folgendem Beitrag soll deshalb gezeigt werden, wie sich solche Gewichtungen einfach berechnen lassen. Als Programmiersprache wurde hierfür die dBASE-Sprache gewählt. Im Gegensatz zu Programmiersprachen wie FORTRAN, C, oder PASCAL, in denen die Programmierung von Ein- und Ausgaberoutinen doch einen recht erheblichen Aufwand erfordert, kann man bei dBASE diese Aufgabe dem Datenbanksystem überlassen. Auf der anderen Seite lassen sich aber auch in dBASE strukturierte, modulare Programme schreiben, die auch zur Lösung von komplexeren Problemen geeignet sind. Und nicht zuletzt zeichnet es sich durch eine große Verbreitung auf PC-Computern aus, sowohl in Bezug auf den Anwender, als auch auf andere Software-Programme, die in der Regel Dateien in dBASE-Format im- und exportieren können. Das folgende Programm läuft unter dBASE III+ und allen dazu kompatiblen Datenbanksystemen (Foxpro, Clipper etc.), mit kleinen Modifikationen unter dBASE III, mit größeren Modifikationen auch unter Version II.

Die Gewichtung durch lokale Trendpolynome

Polynome werden bei der Untersuchung von Zeitreihen angewandt, um den Trend darzustellen bzw. zu eliminieren (8). Solche Polynomanpassungen müssen nicht für die gesamte Reihe berechnet werden, sondern können auch auf einen Stützbereich angewandt werden. Man passt also etwa ein Polynom 3. Grades an die ersten M Werte an und läßt diesen Stützbereich schrittweise über die Zeitreihe gleiten, wobei die Schätzung für den mittleren Wert des Stützbereiches den geglätteten Wert x_t^* ergibt (9). Diese Methode wird z.B. vom Statistischen Bundesamt im Rahmen des Berliner Verfahrens zur Saisonbereinigung angewandt. Die Berechnung der lokalen Trendpolynome ist letztlich nichts anderes, als ein gleitender Mittelwert mit unterschiedlichen Gewichten der einzelnen Elemente. Sie lassen sich -mit einigem Aufwand- für beliebige Polynomgrade und Stützbereiche berechnen. Die für die Praxis sicher ausreichenden Gewichte, die Polynomen 3. und 5. Grades für Stützbereiche von M = 5, 7, 9, ..., 19, 21 entsprechen, sind bei Kendall tabelliert. (10) Um sie auf einen Blick vergleichen zu können, sind sie in den Abb. 7 bis 15 wiedergegeben.

Die Gewichte weisen einen parabelförmigen Verlauf auf, wobei die Parabel jedoch nach unten »verschoben« ist, so daß die äußeren Gewichte negativ sind. Da sich die Gewichte stets zu Eins summieren, wird der Verlauf der Parabel mit zunehmender Weite des Stützbereiches immer flacher. Gewichte nach Trendpolynomen 5. Grades weisen die in den Abb. 17 bis 24 wiedergegebene Gestalt auf. Auch hier werden die Gewichte nach außen hin negativ; anschließend dreht sich das Vorzeichen jedoch noch einmal um. Mit zunehmendem Stützbereich erkennt man immer deutlicher, daß der Verlauf von einem Hochpunkt (t = 0) und zwei Tiefpunkten bestimmt ist. Betrachtet man die Amplitudendiagramme in den Abb. 25 bis 30, wird die Auswirkung dieser Gewichtung deutlich. Bei allen Gewichten sieht man im Vergleich zu den ungewogenen Mittelwerten, daß das Amplitudendiagramm erst später der ersten Nullstelle zustrebt, also mehr Frequenzen in die gefilterte Reihe übertragen werden. Bei der Gewichtung nach einem Trendpolynom 3. Grades und einem Stützbereich von M = 7 ist die erste Nullstelle bei ω_0 = 1,6116 (= 3,9 Zeiteinheiten), bei einem Polynom 5. Grades und dem gleichen Stützbereich sogar erst bei o_0 = 2,4806 (= 2,5 Zeiteinheiten). Es werden also ganz andere Frequenzen eliminiert bzw. übertragen, als bei einem ungewogenen Mittelwert. Leider weisen jedoch auch diese Mittelwerte keinen glatten Verlauf, sondern die schon von den einfachen Mittelwerten her bekannten side-lobes auf, so daß unter diesem Aspekt keine Verbesserung erreicht wird.

Gewichtung nach der Binomialverteilung

Die BinomialVerteilung, die auf J. Bernoulli zurückgeht, gilt als eine der ältesten Verteilungen der Statistik und wurde bereits 1713 veröffentlicht. Sie gibt die Wahrscheinlichkeit an, mit der bei n Versuchen ein Ereignis genau x-mal eintritt. (11) Die Formel hierfür lautet:

$$f(x|n;\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x} \qquad (1)$$

Betrachtet man z.B. einen Münzwurf, so tritt das Ereignis »Wappen« mit der Wahrscheinlichkeit 0,5 ein. Wiederholt man nun diesen Münzwurf n mal, so gibt die Binomialverteilung die Wahrscheinlichkeit dafür an, daß das Ereignis

»Wappen« genau 1 mal eintritt, genau 2 mal eintritt usw. Für 7 Münzwürfe beträgt die Wahrscheinlichkeit, daß

genau 0 mal Ereignis		**»Wappen« eintritt:**	**0,01563**
"	**1**	**"**	**0,09375**
"	**2**	**"**	**0,23438**
"	**3**	**"**	**0,31250**
"	**4**	**"**	**0,23438**
"	**5**	**"**	**0,09375**
"	**6**	**"**	**0,01563**

Ein Stabdiagramm dieser Wahrscheinlichkeit ergibt das in Abb. 31 dargestellte Bild.

Mit diesen Werten lassen sich nun die Koeffizienten der gleitenden Mittelwerte gewichten. Bei einem sieben-gliedrigen Mittelwert wird also nicht mehr jeder Wert mit 1/7 gewichtet, sondern der Wert x_{t-3} mit 0,01563, x_{t-2} mit 0,09375 usw. Dies sind genau die Gewichte, die Ebeling/Irsigler (1977) bei der Glättung der Kölner Getreidepreise verwandt haben und auch von Bauernfeind (1990) übernommen wurden (12).

Solche binomischen Gewichte haben bei der Anwendung auf Zeitreihen wesentlich bessere Eigenschaften, als ungewichtete gleitende Mittelwerte. Beläßt man die Eintrittwahrscheinlichkeit des Ereignisses bei 0,5 so ist auch ein symmetrischer Verlauf der Wahrscheinlichkeitsfunküon um den Wert x_0 gewährleistet; man kann nun für beliebig breite Stützbereiche Gewichte nach diesem Verfahren berechnen.

Betrachtet man die einzelnen Gewichte genauer, so wird allerdings deutlich, daß sie vom mittleren Wert nach außen sehr schnell abnehmen. Dies zeigt besonders deutlich die graphische Darstellung der Gewichte für verschiedene Stützbereiche in Abb. 32. Die entsprechenden Amplitudendiagramme sind in Abb. 33 dargestellt. Man erkennt einen sehr ähnlichen Verlauf unabhängig von der Wahl des Stützbereiches. Rechnet man den x-Achsenabschnitt auf Zeiteinheiten um, so erreicht das Amplitudendiagramm des Stützbereiches M = 7 bei 2,4 Zeiteinheiten praktisch den Wert Null, bei einem Stützbereich von M = 23 bei etwa 3,7 Zeiteinheiten. Betrachtet man die Stellen, an denen das Amplitudendiagramm den Wert 0,2 hat, so wird dieser bei 5,3 (M = 7) bzw. 8,27 (M = 23) Zeiteinheiten erreicht.

Je größer der Stützbereich wird, desto weniger Frequenzen werden übertragen; auch die unerwünschten Neben-Hochpunkte treten nicht auf. Vergleicht man jedoch die verschiedenen Stützbereiche in Hinblick auf ihre Filtereigenschaften, so ergeben sich zwischen den einzelnen Sützbereichen nur tendenzielle Unterschiede. Man könne natürlich das Amplitudendiagramm durch eine entsprechend hohe Wahl des Stützbereiches beliebig nahe an den Wert 0 heranrücken lassen, doch müßte dies mit jeweils (M - 1) / 2 fehlenden Werten an den beiden Rändern der Zeitreihe »erkauft« werden. (13)

Eine erheblich bessere Gewichtung für breitere Stützbereiche wird durch den sog. Kaiserfilter erreicht, dessen Gewichte für einen Stützbereich von M = 23 bei Metz (1988) abgedruckt sind (14). Wie die Abb. 34 zeigt, nehmen die Gewichte nach außen hin deutlich langsamer ab, als bei der entsprechenden Gewichtung nach der Binomialverteilung.

Die Berechnung der Gewichte für beliebige Stützbereiche ist zwar prinzipiell auch möglich, erweist sich jedoch als recht kompliziert (15). Eine Alternative bietet die Gewichtung nach der Normalverteilung, die bereits 1733 veröffentlicht wurde und als wichtigste statistische Verteilung gilt (16).

Gewichtung nach der Normalverteilung

Für unsere Zwecke bietet sie den Vorteil, daß man mit der Varianz die Gewichtung der einzelnen Werte beeinflussen kann. Die allgemeine Formel lautet:

$$f_n(x|\mu;\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (2)$$

Abb. 35 zeigt die Normalverteilungen mit Mittelwert $\mu = 0$ und Varianzen von $\sigma_1 = 1$, $\sigma_2 = 2$ und $\sigma_3 = 3$ und $\sigma_4 = 4$. Durch entsprechende Wahl der Varianz kann also eine beliebige »Dämpfung« der Glockenkurve erreicht werden. Für $\sigma_4 = 4$ zeigt sich ein dem Kaiser-Filter recht ähnlicher Verlauf (vgl. Abb. 36). Vergleicht man die Amplitudendiagramme (Abb. 37-38), so weist der Kaiser-Filter bei 11 Zeiteinheiten die erste Nullstelle auf und hat anschließend sehr kleine side-lobes; er entspricht damit genau einem 11-gliedrigen einfachen Mittelwert ohne dessen Nachteile. Die Funktion der Gewichte nach der Normalverteilung verläuft dagegen glatt, erreicht dafür aber erst bei einer höheren Frequenz den Wert Null, die der Gewichte nach der Binomialverteilung noch später.

Die Programmierung in dBASE

Bei der Untersuchung von Zeitreihen ist man daran interessiert, Stützbereiche verschiedener Länge anzuwenden, etwa für die Untersuchung von saisonalen Schwankungen, Kitchin-Zyklen (3-4 Jahre), Juglar-Zyklen (6-8 Jahre), Kuznets-Zyklen (20 Jahre) usw. Aus diesem Grund wäre es Wünschenwert, wenn man die Gewichtungsfaktoren jeweils in einer seperaten Datei abspeichern könnte, so daß man bei der Ausführung des Programms nur noch deren Namen angeben muß. Diese Datei kann mit jedem beliebigen Editor, der ASCII-Dateien erzeugen kann, geschrieben werden, und hat etwa für einen Mittelwert mit Gewichten nach einem Polynom 5. Grades und M = 21 die Gestalt von Bild 1.

Am Anfang können (beliebig viele) Zeilen, die mit einem Stern beginnen müssen, einen Kommentar enthalten, der Auskunft über die Gewichte gibt. Dann werden die Gewichte beginnend mit v1 an durchnumeriert. Ab dem Wert der dem mittleren folgt, können für eine symmetrische Gewichtung den weiteren Variablen mit v12 = v10, v13 = v9 usw. die gleichen Werte wie den ersten (N-l)/2 zugeordnet werden. Die Gewichte können wahlweise als Bruch oder als Dezimalzahl angegeben werden. Diese Gewichte-Datei wird anschließend vom eigentlichen Programm gewichte.prg eingelesen. Das Programm gewichte.prg beginnt mit dem Löschen des Bildschirms und der Abschaltung der Programmeldungen (Zeile 1 und 2). Anschließend werden 5 String- und eine numerische Variable definiert (Zeile 3) sowie eine Bildschirmmaske (vgl. Bild 2) erzeugt, in der die entsprechenden Dateien angegeben werden können (Zeile 5-17). Im ersten Feld wird der Name der dBASE-Datei, die die Daten enthält angegeben. Anschließend kann - falls die DBF-Datei solch ein Feld enthält - eine Zeitvariable angegeben werden. Existiert diese nicht kann durch den Eintrag »recnoO« eine fortlaufende Numerierung erzeugt werden. Dann wird die Variable angegeben sowie der Name der (ASCII-) Datei, die die Gewichte enthält. Zuletzt muß angegeben werden, wie die (ASCII-) Ausgabe-Datei heißen soll und wie lang der Stützbereich ist. Nachdem die Variablen einlesen worden sind (Zeile 17), wird in der ersten Schleife die Gewichte-Datei eingelesen (Zeile 18 - 24). Dies erfolgt auf die Weise, daß einfach so viele Variablen als public definiert werden, wie im Stützbereich angegeben wurde. Anschließend wird die Gewichte-Datei als Programm ausgeführt, das lediglich die Variablen des Stützbereiches definiert (Zeile 24). Dann wird die Eingabedatei geöffnet und die Ergebnisausgabe in der angegebenen Ausgabedatei protokolliert (Zeile 25 - 27).

Die zweite Schleife (Zeile 28 - 35) schreibt in die Ergebnisdatei für die ersten (M-l)/2 Fälle die Identifikationsvariable, den Originalwert und dann zwei Spalten mit jeweils einer Null für den geglätteten Wert und die Differenz, die für die ersten (M-l)/2 Werte ja nicht berechnet werden können. (17)

In Zeile 36 bis 50 werden dann die eigentlichen Berechnungen durchgeführt. Die äußere Schleife beginnt bei dem ersten Wert nach den (M-l)/2, für die keine Werte berechnet werden können und wiederholt die Berechnung für jeden Fall bis zum rechten Rand, also bis N - (M-l)/2. Die innere Schleife berechnet für jeden Wert den geglätteten als Produkt der M umliegenden mit den jeweiligen Gewichten (Zeile 40 - 44).

Für den ersten Wert einer Zeitreihe sieht dieser Vorgang bei der Berechnung eines sieben-gliedrigen Mittelwertes zum Beispiel so aus:

$$\begin{array}{cccccc} & x_1 & \cdot & v1 & = & wert_1 \\ + & x_2 & \cdot & v2 & = & wert_2 \\ + & x_3 & \cdot & v3 & = & wert_3 \\ + & x_4 & \cdot & v4 & = & wert_4 \end{array}$$

$$
\begin{array}{lllll}
+ & x_5 & \cdot & v5 & = & wert_5 \\
+ & x_6 & \cdot & v6 & = & wert_6 \\
+ & x_7 & \cdot & v7 & = & wert_7 \\
\hline
 & & & \Sigma & & x_4^{G}
\end{array}
$$

Der so berechnete Wert wird in der selben Zeile wie der Originalwert zusammen mit der Differenz dieser beiden Werte ausgegeben (Zeile 45 bis 48).

Schließlich wird der Datensatzzeiger eine Position nach unten versetzt (Zeile 49) und die Berechnung wird von neuem gestartet - so lange bis der Wert N - (M-l)/2 erreicht ist.

Zum Schluß werden die Zeilen 28 - 34 einfach für den rechten Rand der Zeitreihe wiederholt. Die Ergebnisdatei hat für den binomischen 7-gliedrigen Mittelwert die in Bild 4 gezeigte Gestalt (fiktive Zeitreihe, 35 Werte).

Abschließend soll noch kurz gezeigt werden, wie sich für beliebige Stützbereiche Gewichte-Dateien mit Gewichten nach der Binomial- und Normalverteilung automatisch erstellen lassen. Die Binomialverteilung läßt sich am einfachsten nach folgender Rekursionsformel berechnen:

$$f_B(x+1)=f_B(x)\cdot\frac{(n-x)\cdot\theta}{(x+1)\cdot(1-\theta)} \qquad (3)$$

Für die Programmierung sind lediglich drei Schritte notwendig, (vgl. Bild 5). Zuerst wird angegeben, daß die Eintritt-Wahrscheinlichkeit für x = 0 bis N = 6 berechnet werden und eine Ereignis-Wahrscheinlichkeit von $\Theta = 0{,}5$ angenommen wird (Zeile 11 bis 13). Dann wird nach Formel (1) der Wert für x = 0 berechnet: in diesem Fall $f(x) = (1 - \Theta)$ (Zeile 17). Anschließend wird in einer Schleife für jeden weiteren x-Wert nach der Rekursionsformel (3) die entsprechende Wahrscheinlichkeit berechnet und ebenfalls auf dem Bildschirm ausgegeben (Zeile 25 - 28). Als Ergebnis erhält man die in Bild 6 dargestellte Datei.

Die Berechnung der Normalverteilung gestaltet sich nun genau so einfach wie die der Binomialverteilung. Im Gegensatz zu letzterer gebraucht man hierfür jedoch keine Rekursionsformel, sondern berechnet sie direkt durch Angabe der Funktion. Zusätzlich zum Stützbereich kann hier noch die Standardabweichung variiert werden (Zeile 4, Bild 7). Da die Normalverteilung stetig ist und somit für jeden beliebigen x-Wert ein entsprechender y-Wert berechnet werden kann, müssen sich die Gewichte nicht mehr notwendig zu Eins summieren. Mit der Variable summe, die ebenfalls als Kommentar in die Ergebnisdatei übertragen wird (Zeile 24, 26, 27, Bild 7), können die Gewichte daraufhin kontrolliert werden. Für die Eingabe von sigma = 4 und Stützbereich = 23 erhält man z.B. die in Bild 8 dargestellte Gewichtedatei.

Es bleibt nun dem Anwender überlassen, diejenige Gewichtung vorzunehmen, die seinen substanzwissenschaftlichen Vorstellungen am ehesten entspricht. Allgemein läßt sich zu den polynomialen Gewichten sagen, daß sie in der Regel angewendet werden, um Schwankungen zu eliminieren (Ziel II). Abb. 25 bis 30 zeigen allerdings, daß sie dieses Ziel nur unvollkommen erfüllen. Die Gewichte nach der Binomial- bzw. Normalverteilung haben dagegen wesentlich glattere Eigenschaften. Deshalb eignen sie sich eher zur Glättung einer Zeitreihe, um die zyklischen Schwankungen von den zufälligen zu trennen (Ziel I). Durch entsprechende Wahl des Stützbereiches können ganz unterschiedliche Komponenten einer Zeitreihe auf diese Weise untersucht werden.

Vergleicht man solche gewogenen Mittelwerte mit ungewogenen, so sind sie letzteren auf jeden Fall vorzuziehen. Man sollte sich allerdings darüber klar sein, daß z.B. ein gewogener 23-gliedriger Mittelwert (also etwa die hier betrachtete Variante des Kaiser-Filters) in seinem Verhalten mit einem ungewogenen 11-gliedrigen Mittelwert zu vergleichen ist, ein gewogener 13-gliedriger mit einem ungewogenen 7-gliedrigen etc.

Prinzipiell lassen sich auch völlig andere Gewichte als Gewichte-Datei erstellen: etwa die des sog. Parzen-Fensters, oder auch asymmetrische Gewichte. (18) Hilfreich ist auf jeden Fall ein Blick auf das Amplitudendiagramm, um eine erste Vorstellung zu bekommen, welche Auswirkung die Gewichtung hat.

Anmerkungen:

(1) Vgl. Ohler, Norbert, Quantitative Methoden für Historiker. Eine Einführung, München 1980, S. 146ff; Floud, Roderick, Einführung in quantitative Methoden für Historiker, Stuttgart 1980, S. 125ff. Die graphische Darstellung bei Ohler, S. 149 ist irreführend, da er nicht, wie üblich, die Werte um einen symmetrischen Stützbereich gemittelt, sondern dem jeweils letzten Wert zugeordnet hat. Einleuchtender ist die Abbildung bei Floud, S. 126.

(2) Vgl. Weichselberger, K., Über eine Theorie der gleitenden Durchschnitte und verschiedene Anwendungen dieser Theorie, in: Metrika 8 (1964), S. 186.

(3) Der Stützbereich ist die Anzahl der benachbarten Punkte, die in die Mittelwertberechnung eingehen.

(4) Zu dem sog. »Slutzky-Effekt« ist zu bemerken, daß diese künstlichen Zyklen nicht in der geglätteten Reihe auftreten, sondern in den Werten, die aus der Subtraktion von der Originalreihe und dem gleitenden Mittelwert entstehen.

(5) Vgl. als Erläuterung der zugrundeliegenden Ideen Metz, Rainer, Ansätze, Begriffe und Verfahren der Analyse ökonomischer Zeitreihen, in: HSR 47/3 (1988), S. 41ff; zur Berechnung Härtung, Joachim / Elpelt, Bärbel /

Klösener, Karl-Heinz, Statistik. Lehr und Handbuch der angewandten Statistik, München / Wien [7]1989, S. 705ff. Der Terminus ist nicht einheitlich. Metz, S. 64, verwendet die Bezeichnung »Übertragungsfunktion«. Auch der Ausdruck »Frequenz-Antwortfunktion« ist gelegentlich anzutreffen. Der quadrierte Betrag wird in der Regel als »Transferfunktion« bezeichnet. Vgl. Stier, Winfried, Konstruktion und Einsatz von Digitalfiltern zur Analyse und Prognose ökonomischer Zeitreihen, Opladen 1978 (= Forschungsberichte des Landes Nordrhein-Westfalen), S. 3.

(6) Vgl. die Abbildungen bei Metz, Ansätze, S. 64.

(7) Die entsprechenden Verfahren sind beschrieben bei Stier, Konstruktion und Schmidt, R., Konstruktion von Digitalfiltern und ihre Verwendung bei der Analyse ökonomischer Zeitreihen, Bochum 1984.

(8) Vgl. Floud, S. 107ff; Ohler, S. 97ff. Kritisch zu diesem Verfahren Metz, Ansätze, S. 26ff.

(9) Ausführlich zu dieser Methode Leiner, Bernd, Einführung in die Zeitreihenanalyse, München / Wien [2]1986, S. 8 - 53.

(10) Vgl. Kendall, Maurice, Time Series, London [2]1976, S. 31f.

(11) Vgl. etwa Thome, Helmut, Grundkurs Statistik für Historiker. Te il II: Induktive Statistik und Regressionsanalyse, HSR Supplement 3 (1990), S. 37ff.

(12) Vgl. Ebeling, Dietrich / Irsigler, Franz, Getreideumsatz, Getreide- und Brotpreise in Köln 1368-1797. Zweiter Teil: Brotgewichte und Brotpreise: Wochen-, Monats- und Jahrestabelle. Graphiken, Kön / Wien 1977 (= Mitteilungen aus dem Stadtarchiv von Köln 66), S. XIV; Bauernfeind, Walter, Brotgetreidepreise in Nürnberg 1427 - 1538, in: Endres, Rudolf (Hg.), Nürnberg und Bern: zwei Reichsstädte und ihr Landgebiet, Erlangen 1990 (= Erlanger Forschungen A 46), S. 169 - 226.

(13) Aus diesem Grund sind gleitende Mittelwerte für die Untersuchung von langen Wellen oder des Trends ungeeignet.

(14) Vgl. Metz, Ansätze, S. 69.

(15) Die Formel hierfür findet sich z.B. in Stier, Konstruktion, S. 14. Vgl. z.B. zur Berechnung der Besselfunktion etwa Martensen, Erich, Analysis III, Mannheim / Wien / Zürich [3]1986, S. 71.

(16) Vgl. etwa Thome, Statistik II, S. 43ff.

(17) Auf die Problematik des Randausgleichs wird hier nicht näher eingegangen.

(18) Vgl. dazu Weichselberger.

Literaturverzeichnis:

Bauernfeind, Walter, Brotgetreidepreise in Nürnberg 1427 - 1538, in: Endres, Rudolf (Hg.), Nürnberg und Bern: zwei Reichsstädte und ihr Landgebiet, Erlangen 1990 (= Erlanger Forschungen A 46), S. 169 - 226.

Ebeling, Dietrich / Irsigler, Franz, Getreideumsatz, Getreide- und Brotpreise in Köln 1368-1797. Zweiter Teil: Brotgewichte und Brotpreise: Wochen-, Monats- und Jahrestabelle. Graphiken, Köln / Wien 1977 (= Mitteilungen aus dem Stadtarchiv von Köln 66).

Floud, Roderick, Einführung in quantitative Methoden für Historiker, Stuttgart 1980.

Freese, Peter / Müllmerstadt, Friedrich, Standardsoftware Datenbanken: dBASE. Eine strukturierte Einführung, Reinbeck 1988 (= rororo Computer).

Härtung, Joachim / Elpelt, Bärbel / Klösener, Karl-Heinz, Statistik. Lehr und Handbuch der angewandten Statistik, München / Wien [7]1989.

Kendali, Maurice, Time Series, London [2]1976, S. 31.

Leiner, Bernd, Einführung in die Zeitreihenanalyse, München / Wien [2]1986.

Külp, Bernhard u.a., Einführung in die Programmiertechnik ökonomischer Simulationsmodelle in dBASE, Freiburg i. Br. 1988.

Martensen, Erich, Analysis III, Mannheim / Wien / Zürich [3]1986.

Metz, Rainer, Ansätze, Begriffe und Verfahren der Analyse ökonomischer Zeitreihen, in: HSR 47/3 (1988), S. 21 - 103.

Ohler, Norbert, Quantitative Methoden für Historiker. Eine Einführung, München 1980.

Schmidt, R., Konstruktion von Digitalfiltern und ihre Verwendung bei der Analyse ökonomischer Zeitreihen, Bochum 1984.

Stier, Winfried, Konstruktion und Einsatz von Digitalfiltern zur Analyse und Prognose ökonomischer Zeitreihen, Opladen 1978 (= Forschungsberichte des Landes Nordrhein-Westfalen).

Thome, Helmut, Grundkurs Statistik für Historiker. Teil II: Induktive Statistik und Regressionsanalyse, HSR Supplement 3 (1990).

Weichselberger, K., Über eine Theorie der gleitenden Durchschnitte und verschiedene Anwendungen dieser Theorie, in: Metrika 8 (1964), S. 185 - 230.

Historical Social Research, Vol. 18 — 1993 — No. 1, 122-144

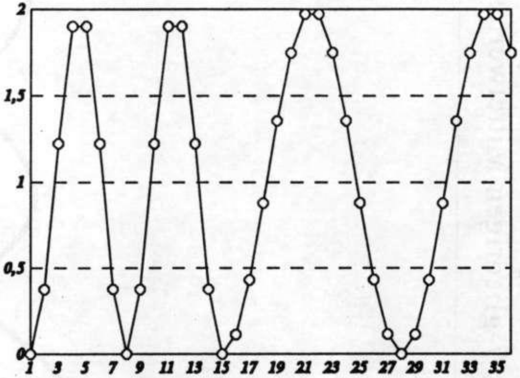


Abb. 1: Originalreihe und ...

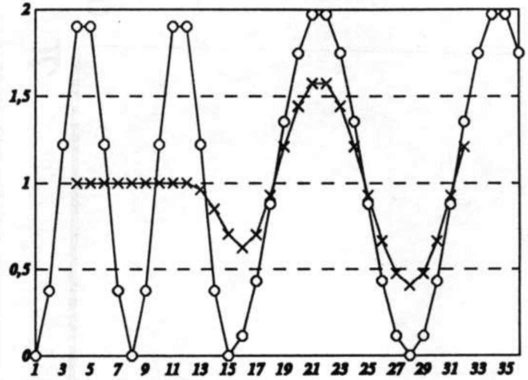


Abb. 2: 7-gliedriger Mittelwert

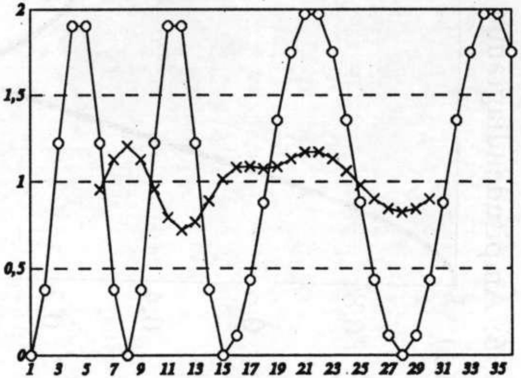


Abb. 3: 11-gliedriger Mittelwert

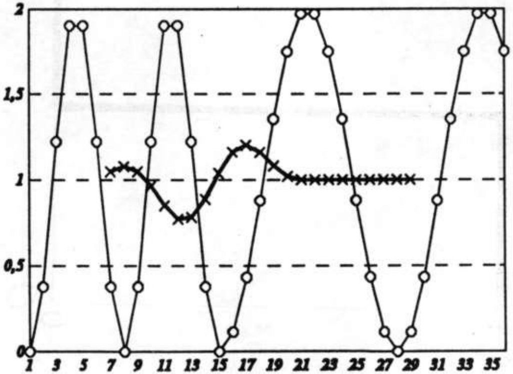


Abb. 4: 13-gliedriger Mittelwert

Abb.5: Ideales Amplitudendiagramm TP7

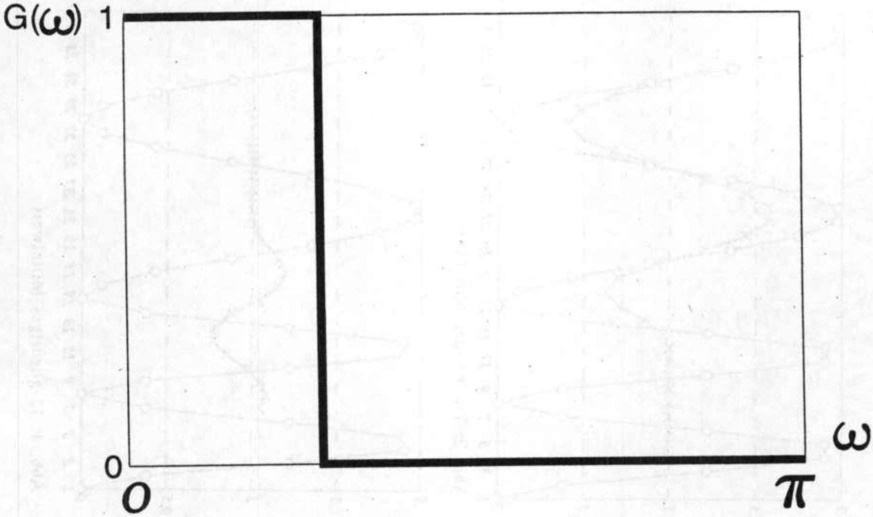
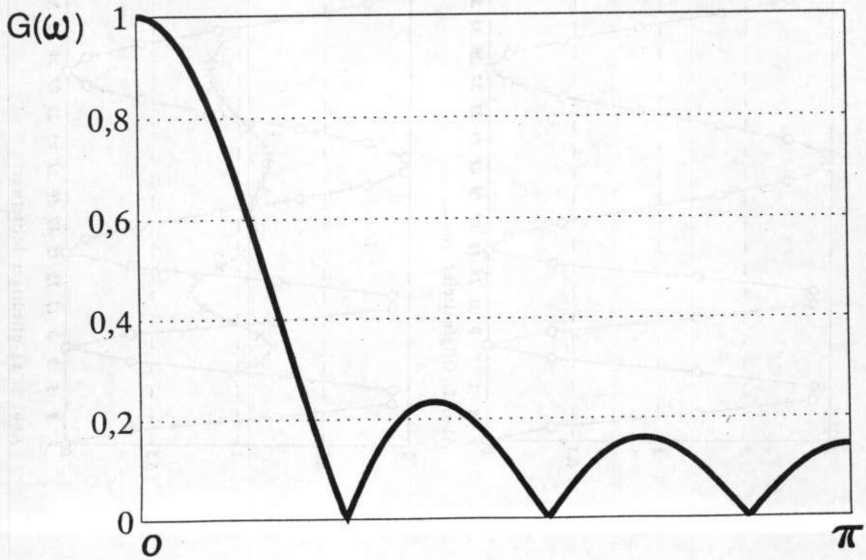


Abb. 6: Ampitudendiagramm des 7-gliedrigen Mittelwertes



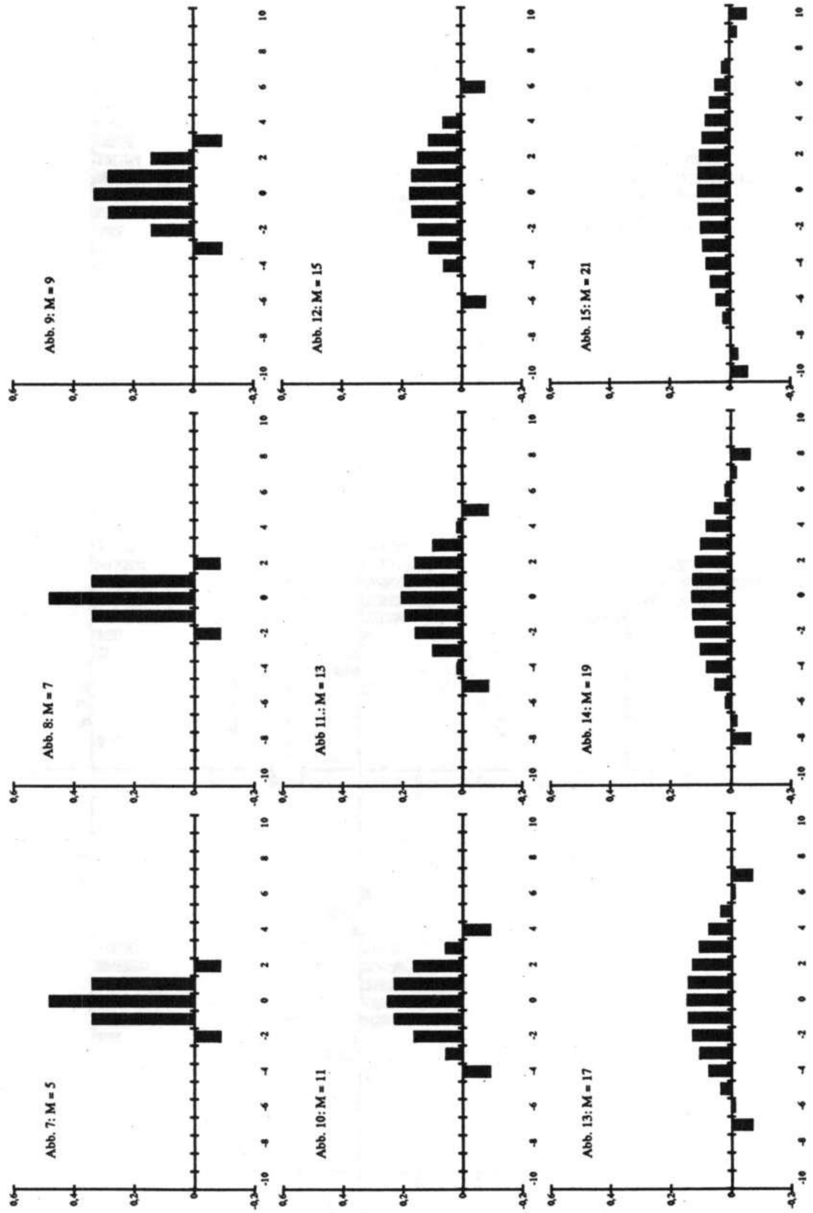


Abb. 7: M = 5
Abb. 8: M = 7
Abb. 9: M = 9
Abb. 10: M = 11
Abb 11.: M = 13
Abb. 12: M = 15
Abb. 13: M = 17
Abb. 14: M = 19
Abb. 15: M = 21

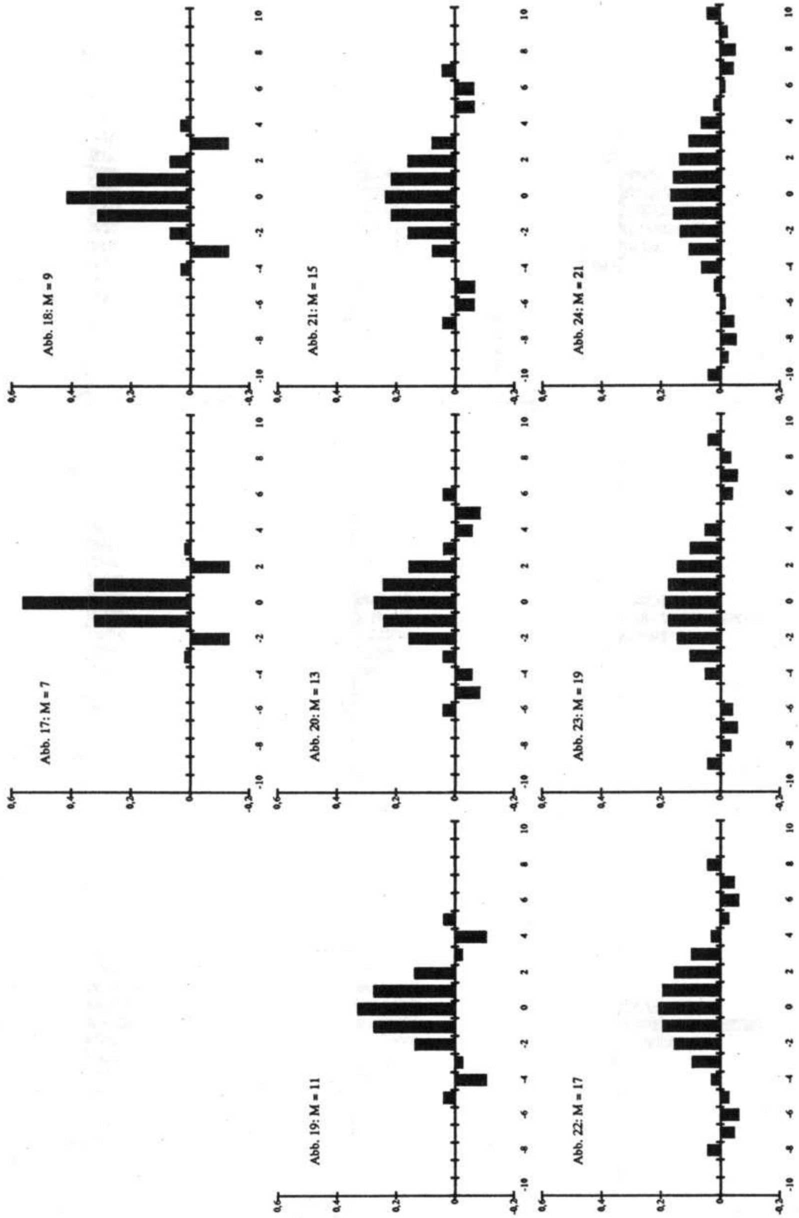
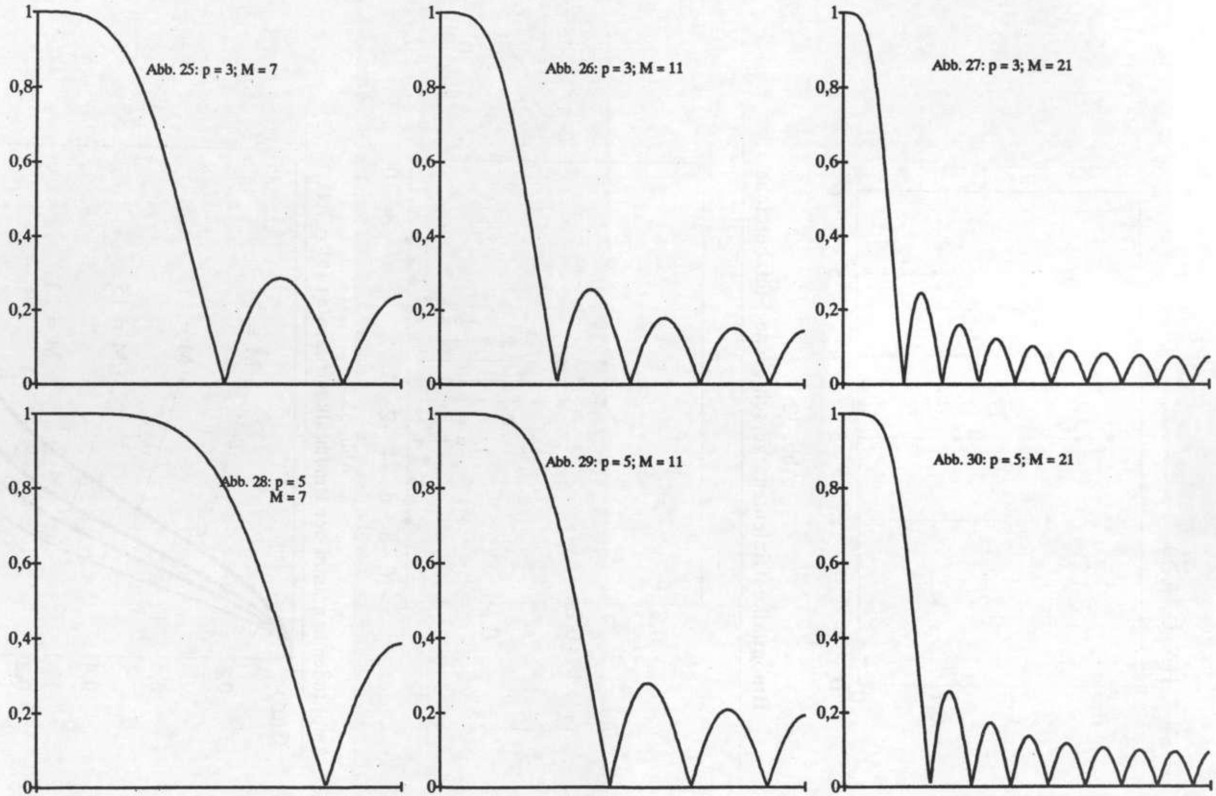


Abb. 17: M = 7
Abb. 18: M = 9
Abb. 19: M = 11
Abb. 20: M = 13
Abb. 21: M = 15
Abb. 22: M = 17
Abb. 23: M = 19
Abb. 24: M = 21

Historical Social Research, Vol. 18 — 1993 — No. 1, 122-144



Historical Social Research, Vol. 18 — 1993 — No. 1, 122-144

Abb. 31: Binomialverteilung

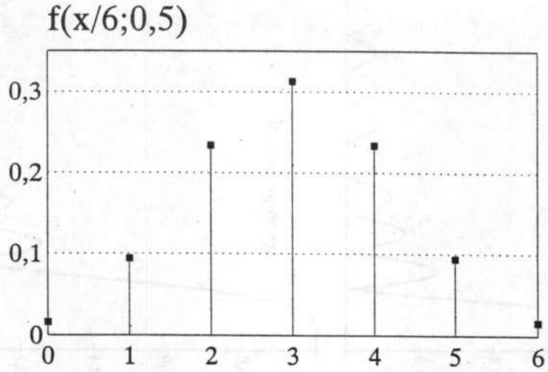


Abb. 32:
Binomialkoeffizienten: Verschiedene Stützbereiche

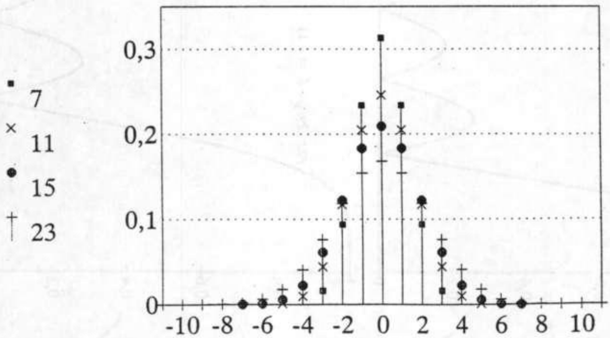


Abb. 33:
Amplitudendiagramm der Binomialkoeffizienten (M = 7,11,15,23)

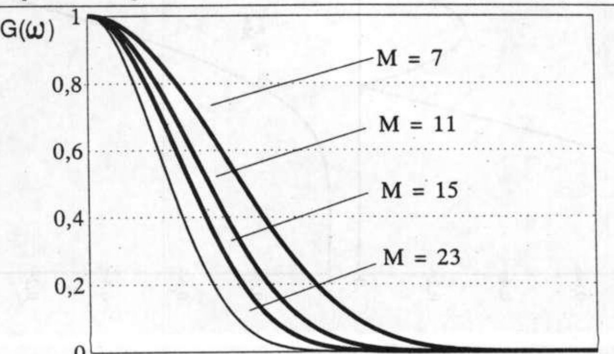


Abb. 34: Gewichtungsfaktoren
Binomial- und Kaiser-Koeffizienten

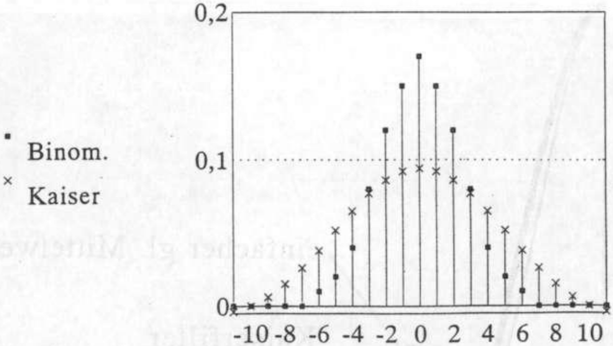


Abb. 35: Normalverteilungsdichten:
Varianzen 1, 4, 9, 16

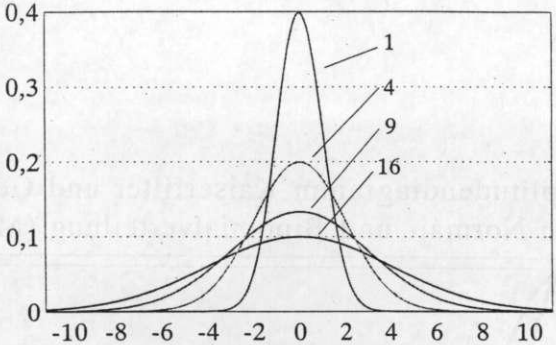
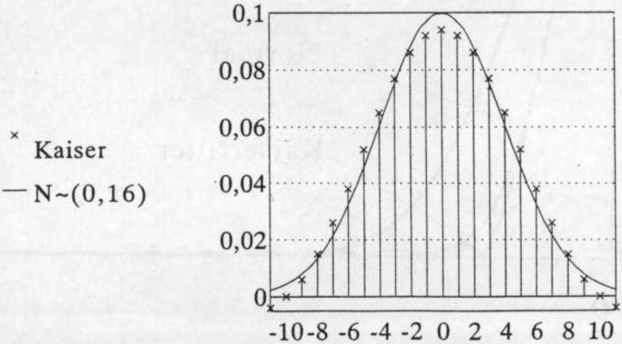


Abb. 36: Gewichtungsfaktoren:
Normal- und Kaiser-Koeffizienten



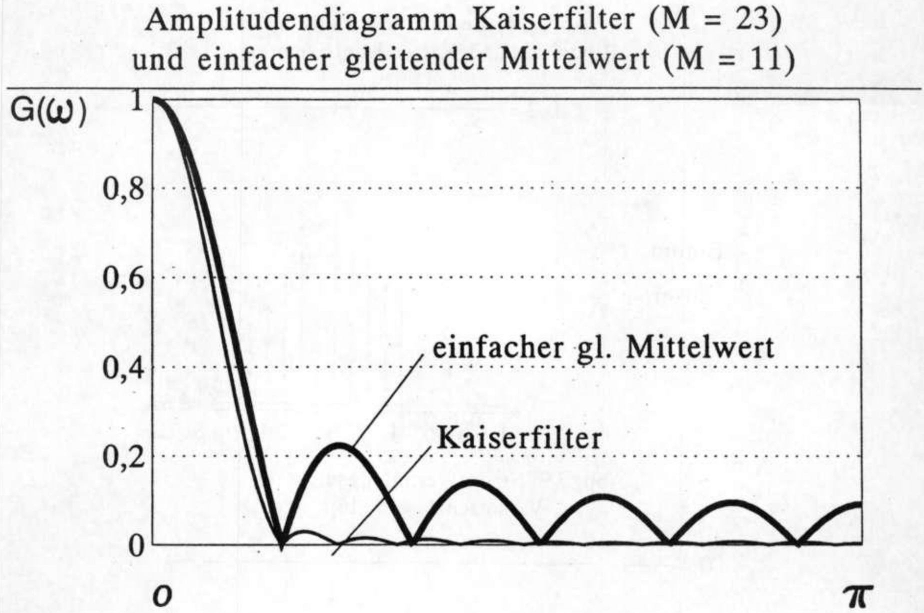

Amplitudendiagramm Kaiserfilter (M = 23)
und einfacher gleitender Mittelwert (M = 11)
G(ω)
1
0,8
0,6
0,4
0,2
0
einfacher gl. Mittelwert
Kaiserfilter
o
π

Abb. 37:

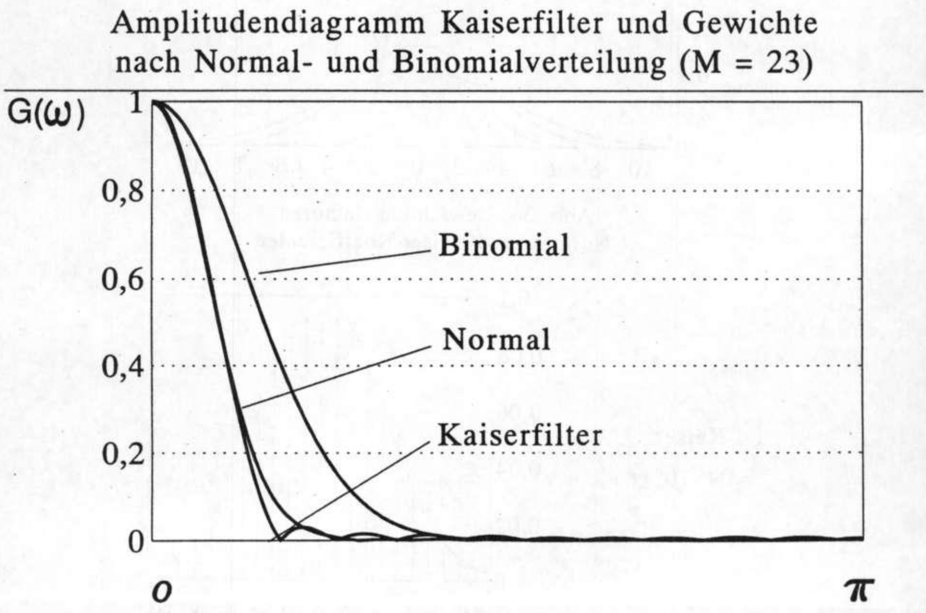

Amplitudendiagramm Kaiserfilter und Gewichte
nach Normal- und Binomialverteilung (M = 23)
G(ω)
1
0,8
0,6
0,4
0,2
0
Binomial
Normal
Kaiserfilter
o
π

Abb. 38:

```
*
*    Polynom 5. Grades, 21-gliedriger Mittelwert
*
v1  =  11628/260015
v2  =  -6460/260015
v3  = -13005/260015
v4  = -11220/260015
v5  =  -3940/260015
v6  =   6378/260015
v7  =  17655/260015
v8  =  28190/260015
v9  =  36660/260015
v10 =  42120/260015
v11 =  44003/260015
v12 = v10
v13 = v9
v14 = v8
v15 = v7
v16 = v6
v17 = v5
v18 = v4
v19 = v3
v20 = v2
v21 = v1
```

Bild 1: Gewichte-Datei `poly5-21.gew`

```
Eingabedatei:

Identifikationsvariable:

Variable:

Gewichtedatei:

Ausgabedatei:

Stützbereich:
```

Bild 2: Eingabebildschirm von Programm `gewichte.prg`

```
clear
set talk off
store space(30) to ident, variable, eingabedatei,
                   gewichtedatei, ausgabedatei
store 0 to stützbereich
@  2,1 SAY "Eingabedatei:"
@  3,1 GET eingabedatei
@  5,1 SAY "Identifikationsvariable:"
@  6,1 GET ident
@  8,1 SAY "Variable:"
@  9,1 GET variable
@ 11,1 SAY "Gewichtedatei:"
@ 12,1 GET gewichtedatei
@ 14,1 SAY "Ausgabedatei: "
@ 15,1 GET ausgabedatei
@ 17,1 SAY "Stützbereich:" PICTURE "@ZZZZ"
@ 18,1 GET stützbereich
read
zähler = 1
do while zähler <= stützbereich
    varnummer = ltrim(str(zähler))
    public v&varnummer
    zähler = zähler + 1
enddo
do &gewichtedatei
use &eingabedatei
set alternate to &ausgabedatei
set alternate on
zähler = 1
do while zähler <= (stützbereich - 1) / 2
   ?? str(&ident,10,5) + "   " + str(&variable,10,5)
   ?? "      0.00000       0.00000"
   ?
   zähler = zähler + 1
   skip
enddo
do while recno()<(reccount()-((stützbereich-1)/2))
   wert = 0
   skip - (stützbereich - 1) / 2
   zähler = "1"
   do while val(zähler) <= stützbereich
         wert = wert + v&zähler * &variable
         skip
         zähler = ltrim(str(val(zähler) + 1))
   enddo
   skip - (stützbereich + 1) / 2
   ?? str(&ident,10,5)+"   "+str(&variable,10,5)+"   "
   ?? str(wert,10,5)+ "   "+str(&variable-wert,11,5)
   ?
   skip
enddo
zähler = 1
do while zähler <= (stützbereich - 1) / 2
   ?? str(&ident,10,5)+"   "+str(&variable,10,5)
   ?? "      0.00000       0.00000"
   ?
   zähler = zähler + 1
   skip
enddo
quit
```

Bild 3: gewichte.prg

1,00000	0,47040	0.00000	0.00000
2,00000	0,79910	0.00000	0.00000
3,00000	0,88330	0.00000	0.00000
4,00000	0,40770	0,72951	-0,32181
5,00000	0,95840	0,76186	0,19654
6,00000	0,79860	0,75982	0,03878
.	.	.	.
.	.	.	.
.	.	.	.
31,00000	0,68860	0,43136	0,25724
32,00000	0,07680	0,39847	-0,32167
33,00000	0,60190	0.00000	0.00000
34,00000	0,24370	0.00000	0.00000
35,00000	0,46100	0.00000	0.00000

Bild 4: Ergebnisdatei

```
set talk off
set echo off
store 0 to stützbereich
clear
@ 5,1 SAY "Wie breit soll der Stützbereich sein ?:"
@ 6,1 GET stützbereich
read
store "binom"+ltrim(str(stützbereich))+".gew" to dateiname
set alternate to &dateiname
set alternate on
x       = 0
N       = stützbereich - 1
theta   = 0.5
?? "* Binomisch gewichteter Gleitmittelwert"
? "* Stützbereich M = "
?? ltrim(str(stützbereich))
f_x = (1 - theta)^N
? "v"
?? ltrim(str(x+1))
?? " = "
?? str(f_x,7,5)
do while x < N
    f_x = (f_x * (N - x) * theta) / ((x + 1) * (1 - theta))
    x = x + 1
    ? "v"
    ?? ltrim(str(x+1))
    ?? " = "
    ?? str(f_x,7,5)
enddo
?
quit
```

Bild 5: binom.prg

```
* Binomisch gewichteter Gleitmittelwert
* Stützbereich M = 7
v1 = 0.01563
v2 = 0.09375
v3 = 0.23438
v4 = 0.31250
v5 = 0.23438
v6 = 0.09375
v7 = 0.01563
```

Bild 6: Ergebnis von Bild 5: `binom7.gew`

```
set talk off
set echo off
stützbereich = 0
sigma        = 0
mu           = 0
summe        = 0
clear
@ 5,1 SAY "Wie breit soll der Stützbereich sein ?:"
@ 6,1 GET stützbereich
@ 8,1 SAY "Sigma: ?"
@ 9,1 GET sigma
read
store "normal"+ltrim(str(stützbereich))+".gew" to dateiname
set alternate to &dateiname
set alternate on
x       = - (stützbereich - 1) / 2
?? "* Normalverteilt gewichteter Gleitmittelwert"
? "* Stützbereich M = "
?? ltrim(str(stützbereich))
zähler = "1"
do while x <= (stützbereich - 1) / 2
      f_x=(1/(sigma*sqrt(2*3.1415926)))*exp(-0.5*((x-mu)/sigma)**2)
      ? "v"
      ?? &zähler
      ?? " = " + str(f_x,7,5)
      x = x + 1
      zähler = ltrim(str(val(zähler) + 1))
      summe = summe + f_x
enddo
? "* Summe: "
?? summe
quit
```

Bild 7: Programm `normal.prg`

```
* Normalverteilt gewichteter Gleitmittelwert
* Stützbereich M = 23
v1 = 0.00227
v2 = 0.00438
v3 = 0.00793
v4 = 0.01350
v5 = 0.02157
v6 = 0.03238
v7 = 0.04566
v8 = 0.06049
v9 = 0.07528
v10 = 0.08802
v11 = 0.09667
v12 = 0.09974
v13 = 0.09667
v14 = 0.08802
v15 = 0.07528
v16 = 0.06049
v17 = 0.04566
v18 = 0.03238
v19 = 0.02157
v20 = 0.01350
v21 = 0.00793
v22 = 0.00438
v23 = 0.00227
* Summe:            0.9960546193
```

Bild 8: Gewichtedatei `normal23.gew`